MODELLING SEAWATER QUALITY OF RACH GIA BAY OF VIETNAM, USING SENTINEL-2 IMAGERY PROCESSED IN THE GOOGLE EARTH ENGINE

Nguyen Hong Quang[1*], Vu Anh Tuan[1], Le Thi Thu Hang[1], Nguyen Tran Dien[2], Le Thanh Son[2], Nguyen Nguyen Minh [3]

[1]Vietnam National Space Center (VNSC) – VAST, [2]Institute of Environmental Technology (IET) – VAST,
[3]Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia

ARTICLE INFO

Received: 30/12/2021

Revised: 25/01/2022

Published: 24/02/2022

KEYWORDS

Chemical oxygen demand
Chlorophyll-a
Google Earth Engine
Sentinel-2
Total Suspended Solid

ABSTRACT

Vietnam has a 3260 km coastline where the seawater quality is affected by activities of agriculture, aquaculture, urbanization, industry, and tourism, raising concerns on coastal ecosystem and social-economic. Hence, monitoring the seawater environment near the coastline is urgently needed. This research exploits the abilities of the support vector machine model to extract surface seawater temperature, total suspended solids, chemical oxygen demand and chlorophyll-a, from Sentinel-2 images processed in the Google Earth Engine platform. The support vector machine model was trained and validated using in-situ measured data and chlorophyll-a data from the National Oceanic and Atmospheric Administration. Model evaluations showed a strong agreement between the modelled and the training data with a mean R^2 greater than 0.8. The water quality distributions presented poorer water quality near the Rach Gia city and river mouths but improved offshore. We strongly recommend this study method and Sentinel-2 data for water quality studies.

MÔ HÌNH HÓA CHẤT LƯỢNG NƯỚC BIỂN TẠI VỊNH RẠCH GIÁ CỦA VIỆT NAM, SỬ DỤNG ẢNH VIỄN THÁM SENTINEL-2 ĐƯỢC XỬ LÝ TRÊN NỀN TẢNG GOOGLE EARTH ENGINE

Nguyễn Hồng Quảng[1*], Vũ Anh Tuân[1], Lê Thị Thu Hằng[1], Nguyễn Trần Điện[2], Lê Thanh Sơn[2], Nguyễn Nguyên Minh[3]

[1]Trung tâm Vũ trụ Việt Nam (VSNC) - Viện Hàn lâm Khoa học và Công nghệ Việt Nam,
[2]Viện Công nghệ Môi trường (IET) - Viện Hàn lâm Khoa học và Công nghệ Việt Nam,
[3]Tổ chức Nghiên cứu Công nghiệp và Khoa học Khối Thịnh vượng Chung, Úc

THÔNG TIN BÀI BÁO

Ngày nhận bài: 30/12/2021

Ngày hoàn thiện: 25/01/2022

Ngày đăng: 24/02/2022

TỪ KHÓA

Nhu cầu oxi hóa học
Diệp lục a
Google Earth Engine
Sentinel-2
Tổng chất rắn lơ lửng

TÓM TẮT

Việt Nam có bờ biển dài 3260 km, nơi chất lượng nước biển đang bị ảnh hưởng bởi các hoạt động nông nghiệp, nuôi trồng thủy sản, đô thị hóa, công nghiệp và du lịch, làm dấy lên lo ngại về hệ sinh thái và kinh tế - xã hội ven biển. Vậy nên, việc quan trắc môi trường nước biển ven bờ là rất cần thiết. Nghiên cứu này khai thác khả năng của mô hình học máy vector hỗ trợ để trích xuất nhiệt độ bề mặt nước biển, tổng chất rắn lơ lửng, nhu cầu oxy hóa học và chất diệp lục-a từ ảnh vệ tinh Sentinel-2, được xử lý trong nền tảng đám mây Google Earth Engine. Mô hình máy vector hỗ trợ đã được đào tạo và kiểm định sử dụng dữ liệu trạm đo và từ Cơ quan Khí quyển và Đại dương Quốc gia Hoa Kỳ. Đánh giá độ chính xác mô hình cho thấy có sự tương quan cao giữa dữ liệu thực đo và dữ liệu mô hình hóa khi giá trị R^2 trung bình lớn hơn 0,8. Bản đồ phân bố chất lượng nước cho thấy chất lượng nước kém hơn ở khu vực gần thành phố Rạch Giá và các cửa sông, và chất lượng nước tốt hơn ở ngoài khơi. Chúng tôi khuyến nghị sử dụng phương pháp nghiên cứu này và ảnh vệ tinh Sentinel-2 cho các nghiên cứu về chất lượng nước.

DOI: https://doi.org/10.34238/tnu-jst.5402

[*] Corresponding author. *Email: nhquang@vnsc.org.vn*

1. Introduction

Recently, the world has faced significant challenges in coastal water quality, mostly in negative manner both in chemical and biological systems due to anthropogenic and physical disturbances [1], [2]. There has been increasing human waste discharging to the sea particularly in the coastal-transitional areas where are developing of the urban, industry, agri/aquaculture, tourism. The flux of nitrogen (N) and phosphorus (P) were dumping to the ocean by 2-fold and 3-fold, respectively [1]. Aware of this risk, the United Nations (UN) has aimed their Sustainable Development Goals (SDGs) of clean water and sanitation (SGD6), life below water (SGD14) [3] all linked to the water quality issue. However, water quality monitoring is complex and costly due to the heterogeneous biological and chemical compounds in the water, hence researchers are searching for appropriate approaches such as modelling [4] and remote sensing techniques, particularly for large study areas.

Vietnam has a long coastline of 3260 km [5]. Its coastal water quality could be divided into three distinct zones: of the gulf of Tonkin (mostly affected by the Red River Delta); the central Vietnam sea (from Quang Tri to Ninh Thuan province, affected by industry and tourism); and the Southern Vietnam sea (from Binh Thuan to Ca Mau, driven by discharge from the Mekong river to the sea and tourist activities). In recent years, seawater quality problems in Vietnamese sea have been raised at a higher level due to a series of “environmental scandals” such as the Formosa Ha Tinh waste leak [6], industrial wastewater discharges to the Thi Vai river [7]. That has negatively affected the sea ecosystem environment and the short-term and long-term livelihoods of coastal communities, noting the importance of seawater monitoring. In Rach Gia city of Kien Giang province in the Mekong delta, tourism, fisheries, and fast growth of urbanization could lead to water environmental issues if the management is inadequate. This study focuses on the Rach Gia Bay where discharge water from agriculture, aquaculture, domestic wastewater from urban and tourist waste could alter the water quality. Therefore, the water quality of the Rach Gia Bay is driven by dumping materials from these activities which are expected to increase currently and in the future. Additionally, the main sea transportation from inland to the Phu Quoc Island is through the bay and contributes to the water quality reduction of the bay’s water.

Recently, remote sensing technique has emerged as an effective method in detecting river hydrology and water and sediment dynamics [8], [9]. It can be employed to monitor changes in water quality parameters (total suspended solid (TSS), chlorophyll, and temperature) [10], [11]. It is noted that there are hundreds of water quality parameters, but not all of them can be extracted from remote sensing data. However, in terms of generating quick but reliable results covering very large areas, remote sensing technique has advantages over traditional approaches [12]. In addition, remote sensing data and analytics platforms are increasingly available. Many data are free of charge and available globally, besides some cloud computational platforms like Google Earth Engine (GEE), Amazon Web Services, Data Cubes,... All facilitates extracting and mapping the water quality indicators with less time and cost. This study models some water quality indicators including sea surface temperature (SST), chlorophyll-a (chl-a), total suspended solids (TSS), and chemical oxygen demand (COD) computed in the GEE for Rach Gia Bay from 2015 to 2021. Those first three parameters are commonly found in remote sensing application as they are highly sensitive to visible reflectance of optical remote sensing data. We tested the capability of Sentinel-2 to extract the COD parameter as it could be less responsive to the remote sensing data.

2. Materials and Methods

2.1. Study area and data collection

Rach Gia Bay is quite small, located in the south-west Ca Mau Peninsula (Figure 1). Its center coordinate is 105.01°E, 10.00°N. The bay area is approximately 1,705 km^2, receiving discharges from Cai Lon and Rach Gia-Long Xuyen rivers and the dense channel system connected to the Hau river. Rach Gia city is a small city with a population of 230 thousand people. However, the city has been growing rapidly in recent years in terms of area, population, and economy.

There are two water quality measurement stations (S1 and S2 in Figure 1) in the bay measuring various water quality parameters. Among them, water temperature, total suspended solid (TSS) and chemical oxygen demand (COD) were used for this study. The water quality data of the two stations are available from 2015 to 2020. The water samples were collected and analyzed following the standard (QCVN 10-MT:2015/BTNMT) of the Ministry of Natural Resources and Environment (MONRE) to ensure the accuracy of the results.

The near real-time chlorophyll-a data (level 3) were collected for 2020 to 2021 period, from a data archiving of the National Oceanic and Atmospheric Administration (NOAA). The data are available globally and free of charge on [13].

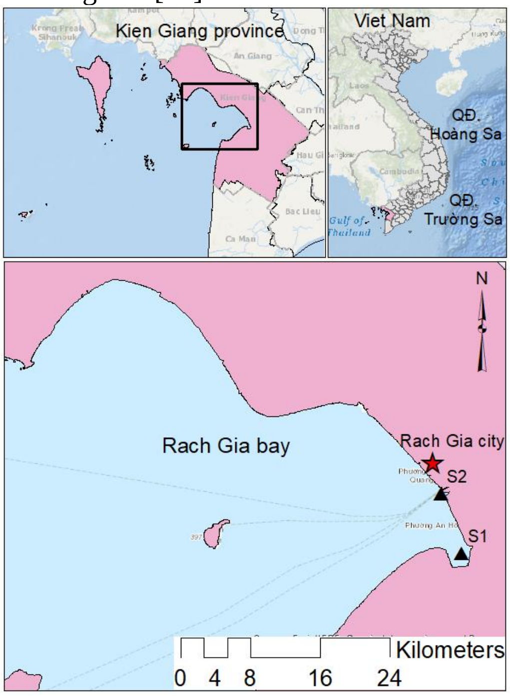

Figure 1. *Location of the Rach Gia Bay*

2.2. Methods

Our study work-flow is depicted in Figure 2. The central process of modelling the water quality indicators (i.e. temperature, total suspended solids, COD and chlorophyll-a) includes supporting region of interest (ROI) extraction and surface reflectance calculation components, using Normalized Difference Water Index (NDWI) and training data. The Sen2cor developed by European Space Agency (ESA) team aims to correct Sentinel-2 Level-1C products from the effects of the atmosphere delivering Level-2A surface reflectance products [14]. Subsequently, the raster-based outputs of the machine models will be further processed to generate appropriate thematic water quality maps. The machine learning models and model accuracy evaluation are presented in more details as follows.

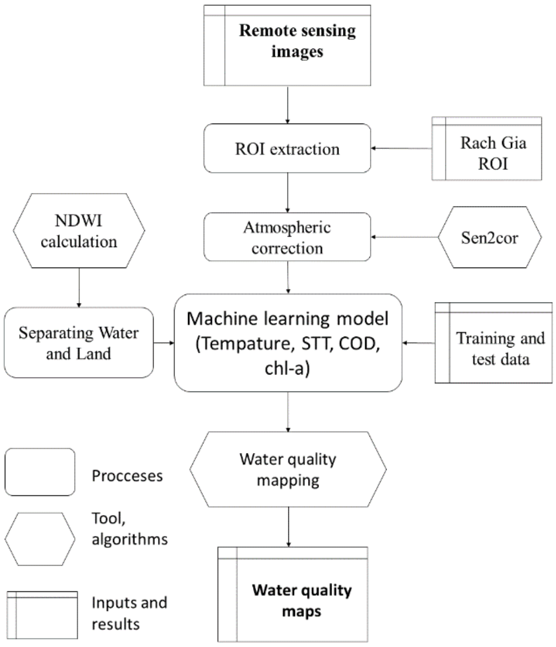

Figure 2. *Processing chain applied to Sentinel-2 images for water quality mapping*

2.2.1. Machine learning models

Machine learning (ML) algorithms have been widely used in image classifications and regression. They accept a variety of training data and are routinely found to have higher accuracies than other unsupervised classifiers [15]. Currently, several ML classifiers including CART, Random Forest, Naïve Bayes and Support Vector Machine (SVM), are integrated into the Earth Engine API and available in Python and JavaScript. This provides many advantages of an online computing platform (i.e. fast, no installation required) and variations of remote sensing data sources (free of charge).

The SVM [16] is a supervised non-parametric statistical learning techniques that could be applied for complex and noise data to generate precise classification results [17]. However, this classifier is often time-consuming to process large time-series images. The SVM statistical learning theory is the system that separates the classes with a decision surface maximizing the margin between the classes. The surface is called the “optimal hyperplane” and the data points nearest the hyperplane are called “support vectors” [18].

For applying SVM in GEE, five procedures have been implemented in this study, including:

- Processing ground-truth data from the S1 and S2 stations (for temperature, TSS and COD), and the chlorophyll-a data downloaded from the NOAA website, containing fields of class labels and values,
- Instantiating the SVM classifier and setting the model parameters,
- Training the SVM classifier using training data (random of 70% of the ground-truth dataset),
- Running the model to classify the images,
- Estimating classification errors with the independent validation of the training data (random of 30% of the ground-truth dataset).

These modelling procedures were applied identically for all water quality parameters of sea surface temperature (SST), total suspended solids (TSS) and chemical oxygen demand (COD) and Chlorophyll-a concentration. A Python 3.8 program was integrated as a computing kernel in a Jupyter notebook to perform this task.

2.2.2. Model accuracy evaluation

We used three common statistical error indicators: the coefficient of determination (R^2) by Equation 1, the root-means-square deviation (RMSE) by Equation 2, and the bias error by Equation 3. The R^2 values approaching 1.0 means the model results are approaching the optimal values (equal to the comparison values). The larger the RMSE and absolute bias values indicate higher uncertainty of the model results.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{y}_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (2)$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \quad (3)$$

Where x_i is model-estimated value, y_i is in-situ measured value and $\overline{y}_i$ is the mean value of the in-situ measured values.

3. Results and Discussion

3.1. Time series of spatial water quality

Figure 3 shows the comparisons between the SVM model-predicted time-series results of four water quality parameters of sea surface temperature (SST), total suspended solids (TSS), chemical oxygen demand (COD) and chlorophyll-a concentration (Chl-a) with the in-situ measurement at S1 and S2 stations and the chl-a information derived from NOAA (August 2015 to November 2021) [13]. The highest agreements are obtained for SST and TSS. However, there was a rough error of estimated SST (with a minus value in January of 2019). As the input image was not downloaded (remaining in GEE), the source of this error remained unknown. The remote sensing derived COD results seemed to be slightly overestimated compared to the in-situ measurement data, particularly for July 2017.

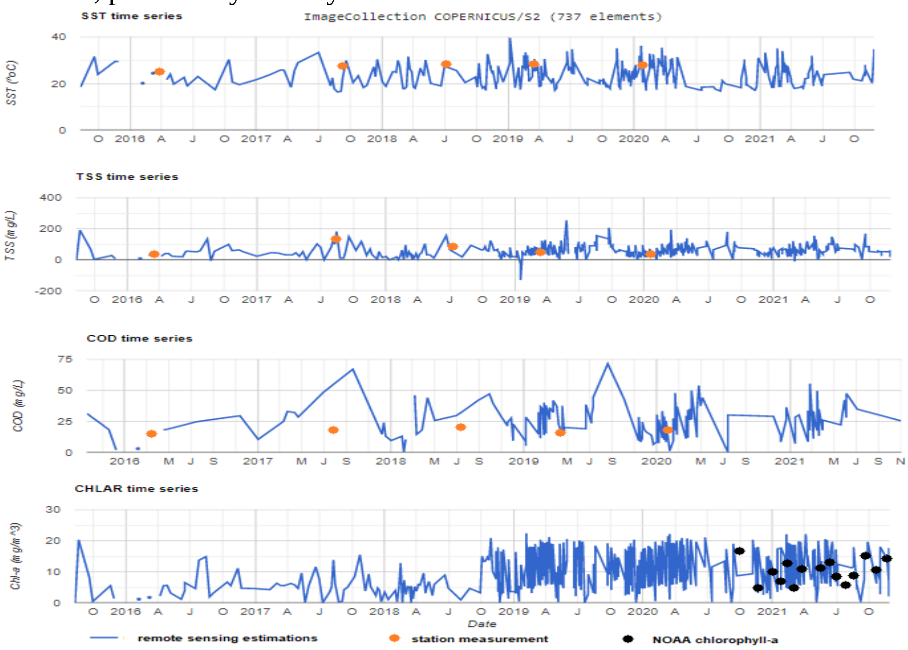

Figure 3. *Time series of sea surface temperature (SST), total suspended solids (TSS), chemical oxygen demand (COD), and chlorophyll-a concentration modelled from Sentinel-2 images from August 2015 to November 2021 in comparison with measured data of stations S1 and S2 and NOAA chlorophyll-a*

COD parameter is an important parameter that indicates organic matter content in water and thus presents how much water is polluted. However, this is quite difficult to estimate COD from remote sensing data. Similar results were found by Sharaf et al 2017 [19], where the COD estimations had larger errors than other parameters such as TSS and turbidity. On the other hand, water chlorophyll-a (mainly indicating algae) estimation is a reliable and, thus, more common remote sensing application. Because the chlorophyll-a is sensitive to the B4 (red) and B5 (vegetation red edge) of Sentinel-2 images [20]. We found good agreement between selected weekly NOAA chlorophyll-a data with our Sentinel-2 based estimates in the period from October 2020 to November 2021. The NOAA oceanic data are a good reference source, particularly for a large study area (available globally) and high temporal resolution data (weekly).

3.2. Model evaluation

In general, the model outputs of the training phase were estimated with higher accuracy than those of the validation phase (Table 2). While the R^2 values indicated the goodness of fit between the model estimated water quality indicators with the measured data from the S1 and S2 stations and NOAA reference data, the bias also showed the model under (negative) or over (positive) estimation of these values. Hereby, the model seemed to over-estimated the SST but under-estimated the TSS, COD and Chl-a (negative bias values) in general. The R^2 values of all parameters revealed a strong agreement of modelled and observed data. However, the R^2 of SST and TSS were higher than R^2 estimated for COD and chl-a. Comparing results from previous studies, the statistical uncertainty indicators of this study could be considered acceptably [19], [21]. The accuracy of the SVM models could be improved if we had a longer and higher temporal resolution of the training dataset. Unfortunately, better chl-a data (gauge data or finer spatial resolution grids) were unavailable or not available for this study. The NOAA data was a relevant alternative.

Table 1. *Statistical uncertainty Indicators*

	R^2	**RMSE**	**Bias**	**R^2**	**RMSE**	**Bias**
Parameters	**Training phase**			**Validation phase**		
SST (°C)	0.92	0.34	0.42	0.91	0.36	0.51
TSS (mg/L)	0.90	0.50	−0.25	0.87	0.53	−0.32
COD (mg/L)	0.71	1.71	−2.62	0.68	2.57	−3.21
Chl-a (mg/m^3)	0.89	1.12	−1.56	0.85	1.65	−2.09

3.3. Water quality mapping

Mean annual water quality of Rach Gia Bay was mapped and depicted in Figure 4 for four parameters of SST, TSS, COD and chl-a. The upper left is the SST map with higher temperature (28-29 degree Celsius) in the areas near the Rach Gia city and the coastline and the SST was lower about one degree Celsius further offshore. Similarly, the spatial distribution of TSS and COD (mg/L) were higher in the Rach Gia Bay with values of around 150 mg/L of TSS and 125 mg/L of COD. These two parameters are strongly linked to the urban waste and river discharge indicating the poorer seawater quality of the bay compared to the further areas in the sea [22], [23]. The chl-a concentrated more on the Cai Lon River mouth (16-20 mg/m^3) but also scattered in the sea with dominant values of 11-15 mg/m^3. There has not been found relevant support values of these water quality indicators from previous studies for comparisons. However, logically, with the effects of domestic wastewater, aquaculture and agriculture discharging to the bay, the higher values of STT, TSS and COD near the coastline of Rach Gia Bay were reasonable. This assumption was supported by Long et al. (2016) [22].

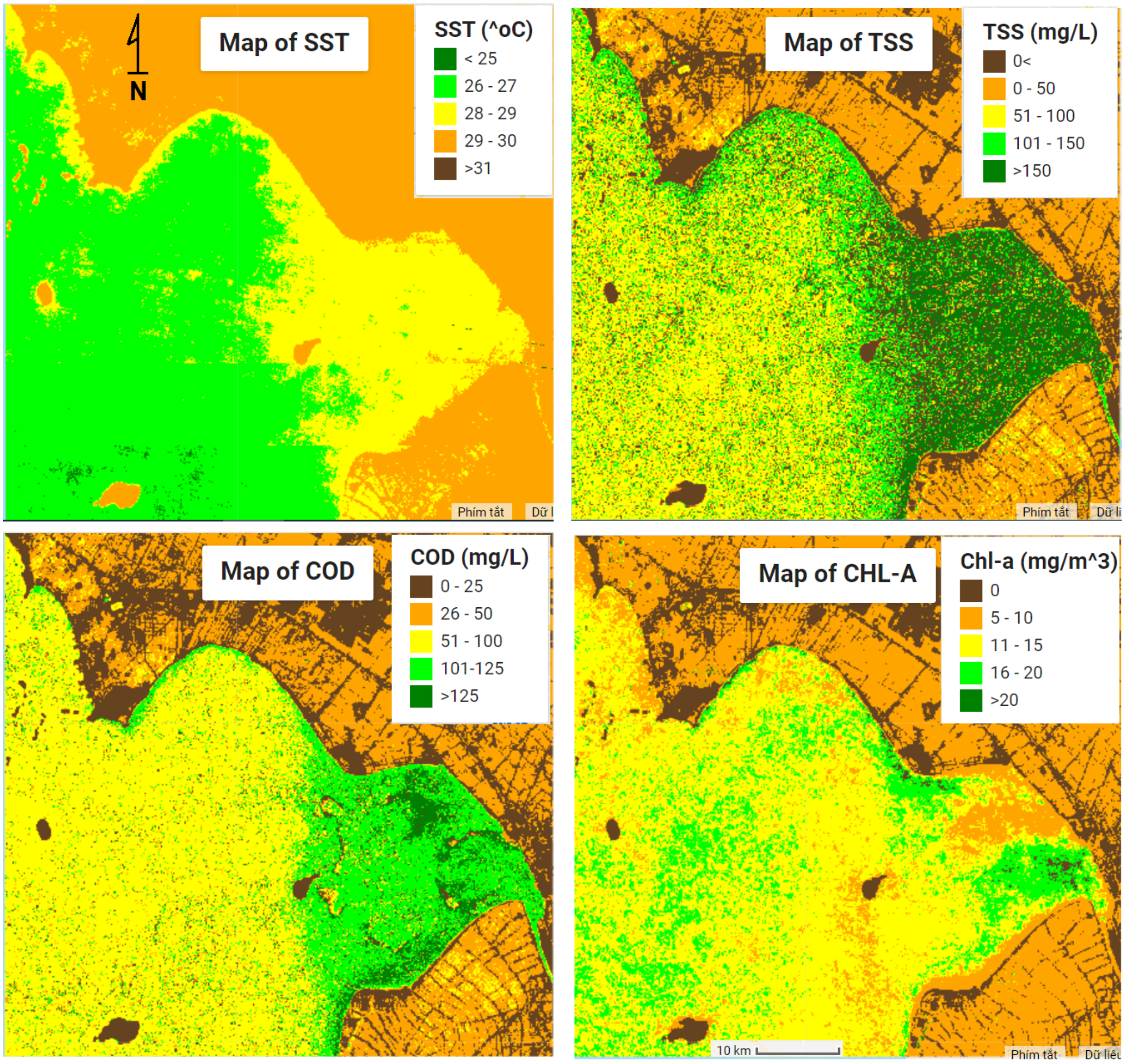

Figure 4. *Maps of average seawater quality parameters of sea surface temperature (SST), total suspended solids (TSS), chemical oxygen demand (COD) and chlorophyll-a (CHL-A)*

4. Conclusion

- Support vector machine (SVM) is a machine learning classifier that is commonly used in remote sensing modelling and classification. It was successfully tested in this study for the seawater quality modelling of Rach Gia Bay using the Sentinel-2 images.
- This study used the Google Earth Engine, which showed to be an effective and convenient cloud computing platform for fast task execution and free of charge. There are no powerful computers required to store and process a large image dataset.
- SVM water quality values were validated using observed data from in-situ measurement stations and NOAA. Good agreements between modelled data and observed data are obtained. Longer and denser on-site measured data used for model train could guaranty a better model outcome. Sentinel-2 can be considered a good data source for water quality parameter extractions.
- Linkages between seawater quality and driven sources were not analyzed. However, it was clearly shown that all the estimated indicator values were higher (i.e. poorer water quality) in the areas near the city and river mouths; and were lower (i.e. better quality) in the sea offshore.

Optical remote sensing technique is limited at its ability to extract some sensitive water quality parameters to reflectance at the visible spectrum. Many other parameters are remaining unable to detect or to over/under-estimate them with high uncertainty like the COD.

Acknowledgments

The authors would like to thank the Ministry of Science and Technology (MOST) of Vietnam for supporting the project “Supporting rapid assessment of landscape change to improve planning and decision-making in Vietnam” (Research grant No. NĐT/AU/21/15).

REFERENCES

[1] R. W. Howarth, A. Sharpley, and D. Walker, "Sources of nutrient pollution to coastal waters in the United States: Implications for achieving coastal water quality goals," *Estuaries,* vol. 25, pp. 656-676, 2002.

[2] M. E. Baird, M. Mongin, J. Skerratt, N. Margvelashvili, S. Tickell, and A. D. Steven, "Impact of catchment-derived nutrients and sediments on marine water quality on the Great Barrier Reef: An application of the eReefs marine modelling system," *Marine Pollution Bulletin,* vol. 167, p. 112297, 2021.

[3] UNDP, "United Nations Development Programme," 2015. [Online]. Available: *https://www.undp.org/content/undp/en/home/sustainable-development-goals.html#:~:text=The%20Sustainable%20Development%20Goals%20(SDGs,peace%20and%20prosperity%20by%202030.* [Accessed March 2021].

[4] I. James, "Modelling pollution dispersion, the ecosystem and water quality in coastal waters: a review," *Environmental Modelling & Software,* vol. 17, pp. 363-385, 2002.

[5] H. Takagi, N. D. Thao, M. Esteban, T. Mikami, and V. T. Ca, "Coastal disasters in Vietnam," in *Handbook of coastal disaster mitigation for engineers and planners*, ed: Elsevier, 2015, pp. 235-255.

[6] M.-F. Fan, C.-M. Chiu, and L. Mabon, "Environmental justice and the politics of pollution: The case of the Formosa Ha Tinh Steel pollution incident in Vietnam," *Environment and Planning E: Nature and Space, doi:10.1177/2514848620973164*, 2020.

[7] K. Prilop, H. Q. Nguyen, M. Lorenz, H. Le, T. H. Le, and G. Meon, "Integrated water quality monitoring of the Thi Vai River: an assessment of historical and current situation," in *EWATEC-COAST: Technologies for Environmental and Water Protection of Coastal Zones in Vietnam. Contributions to 4th International Conference for Environment and Natural Resources, ICENR 2014*, 2014, pp. 97-110.

[8] M. Nones and M. Guerrero, "GIS and remote sensing for tracking morphological changes and vegetation coverage at the reach scale: the Po River case study," *In River Flow 2020.* CRC Press, 2020, pp. 1005-1012.

[9] M. Nones, "Remote sensing and GIS techniques to monitor morphological changes along the middle-lower Vistula river, Poland," *International Journal of River Basin Management,* vol. 19.3, pp. 345-357, 2021.

[10] J. C. Ritchie, P. V. Zimba, and J. H. Everitt, "Remote sensing techniques to assess water quality," *Photogrammetric engineering & remote sensing,* vol. 69, pp. 695-704, 2003.

[11] J. A. Griffith, "Geographic techniques and recent applications of remote sensing to landscape-water quality studies," *Water, Air, and Soil Pollution*, vol. 138, pp. 181-197, 2002.

[12] M. Elhag, I. Gitas, A. Othman, J. Bahrawi, and P. Gikas, "Assessment of Water Quality Parameters Using Temporal Remote Sensing Spectral Reflectance in Arid Environments, Saudi Arabia," *Water*, vol. 11, p. 556, 2019.

[13] NOAA, "National Oceanic and Atmospheric Administration Archive," 2021. [Online]. Available: *https://data.noaa.gov/dataset/dataset/chlorophyll-noaa-viirs-snpp-near-real-time-global-level-3-2018-present-experimental-weekly.* [Accessed October 2021].

[14] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, "Sen2Cor for sentinel-2," in Image and Signal Processing for Remote Sensing XXIII, 2017, *International Society for Optics and Photonics*, pp. 1042704.

[15] H. Q. Nguyen, C. H. Quinn, L. C. Stringer, R. Carrie, C. R. Hackney, T. V. H. Le, V. T. Dao, and T. T. N. Pham "Multi-Decadal Changes in Mangrove Extent, Age and Species in the Red River Estuaries of Viet Nam," *Remote Sensing,* vol. 12, p. 2289, 2020.
[16] V. Vapnik, *The nature of statistical learning theory*. 840 Springer-Verlag New York, Inc., New York, NY, USA, vol. 841, p. 842, 1995.
[17] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research,* vol. 5, pp. 975-1005, 2004.
[18] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, *A practical guide to support vector classification*. Taipei, Taiwan, 2003.
[19] E. S. El Din, Y. Zhang, and A. Suliman, "Mapping concentrations of surface water quality parameters using a novel remote sensing and artificial intelligence framework," *International Journal of Remote Sensing,* vol. 38, pp. 1023-1042, 2017.
[20] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," *Sensors,* vol. 11, pp. 7063-7081, 2011.
[21] G. Sent, B. Biguino, L. Favareto, J. Cruz, C. Sá, and A. I. Dogliotti, "Deriving Water Quality Parameters Using Sentinel-2 Imagery: A Case Study in the Sado Estuary, Portugal," *Remote Sensing,* vol. 13, p. 1043, 2021.
[22] T. L. Trinh, M. C. Dang, T. V. Pham and C. C. Duong, "Impacts of Urban Wastewater on Water Quality of the Lake at Rach Gia Bay in the Mekong Delta, Vietnam," *Journal of Geography, Environment and Earth Science International,* vol. 5, pp. 1-12, 2016.
[23] Q. T. Doan, T. M. N. Nguyen, H. T. Tran, and J. Kandasamy, "Application of 1D–2D coupled modeling in water quality assessment: A case study in Ca Mau Peninsula, Vietnam," *Physics and Chemistry of the Earth, Parts A/B/C,* vol. 113, pp. 83-99, 2019.